Spectral Detection of Human Targets

Dalton Rosario

Signal and Image Processing Division, Army Research Laboratory, 2800 Powder Mill Road

Adelphi, MD, USA

dalton.s.rosario.civ@mail.mil

Abstract- **A three-stage algorithm suite is proposed for a specific human target detection scenario, where a visible/near infrared hyperspectral (HS) sample is assumed to be available as the only cue from a reference image frame. The suite first applies a biometric based human skin detector to focus the attention of the search. Using as reference all of the bands in the spectral cue, the suite follows with a Bayesian Lasso inference stage designed to isolate pixels representing the specific material type cued by the user and worn by the human target (e.g., hat, jacket). In essence, the search focuses on testing material types near skin pixels. The third stage imposes an additional constraint through RGB color quantization and distance metric checking, limiting even further the search for material types in the scene having visible color similar to the target color. Using the proposed cumulative evidence strategy produced some encouraging range-invariant results on real HS imagery, dramatically reducing to zero the false alarm rate on the example dataset. These results were in contrast to the results independently produced by each one of the suite's stages, as the spatial areas of each stage's high false alarm outcome were mutually exclusive in the imagery. These conclusions also apply to results produced by other standard methods, in particular the kernel SVDD (support vector data description) and Matched Filter, as shown in this paper.**

Keywords- Human Target Detection; VNIR; Hyperspectral

I. INTRODUCTION

In the past two decades, hyperspectral (HS) remote sensing has found many applications in both commercial and military communities, in particular, in earth observation, targeting, and intelligence, surveillance, and reconnaissance. HS sensors collect data in several narrow and adjacent spectral bands, thus providing a densely sampled spectrum for each pixel in the scene. Such a high spectral resolution preserves important aspects of the spectrum and makes it possible to reveal subtle spectral characteristics of a given material. In fact, HS sensing has proven valuable for discrimination of materials on the basis of their unique *spectral signature*, which is the spectral reflectance in the visible/near infrared (NVIR) and shortwave infrared (SWIR) regimes [1].

The utility of hyperspectral sensors is certainly not limited to applications of physical sciences and can be extended to an excess of remote object detections such as aircraft wreckage or a lost hiker's exposed skin [2]. In terms of the latter example, current airborne systems exist to perform search and rescue by way of spectral matching [3]. The effectiveness of the rescue application is often in the hands of an operator that requires a large degree of *operator capability*. It is noted by experts that any hyperspectral system developed for use in rescue applications will need to be simple enough to operate by a non-hyperspectral-exploitation expert and allow the user to discriminate small targets in a large scene. Due to the nature of rescue applications, near real-time exploitation of the dataset is essential for it to be of use [4]. A hyperspectral/multi-spectral system designed to automatically detect and classify the pigmentation level of human skin can be one of the components of such a critical system.

Face recognition is another active area of research that could benefit from knowledge of the spectrum of human skin. Many face recognition algorithms depend on the size, shape, and color of facial features to identify a face [5]. However, performance of these algorithms is degraded by something as simple as a change in face orientation [6]. Hyperspectral images provide several additional features to aid in face recognition. For example, the spectral signature of skin on the face provides a pose-invariant feature for use in these algorithms [7].

In this work, I am interested in using VNIR HS remote sensing for the purpose of detecting and possibly tracking an object of interest (target) in any given context (e.g., populated urban area) during a particular time interval. Assume further that the target is a person, who may be unknown a priori by an interested party until the party (given a justifiable reason)determines the person under initial observation to be a target at a particular point in time. Furthermore, the target may be partially or fully obscured at times, for short or long periods, during the particular time interval of interest; and its kinematic state may vary (e.g., stationary, moving at a constant or varying speed). In addition, the viewing perspective of the sensor is not nadir, but either elevated slant or ground to ground.

Some of the challenges imposed by the assumed scenario mentioned above include: (a) the exclusion of advanced classification algorithms from consideration (e.g., support vector machine [8]), since these classifiers require prior knowledge of both background clutter and target for training, whereas in the above scenario the background can potentially vary due to the natural scene dynamics; (b) the exclusion of standard tracking algorithms from consideration (e.g., Kalman Filtering [9]), since the target's relative motion is not a reliable tracking feature during periods of full obscuration, target proximity to other moving objects, and changes to the target's kinematic state; and last but not least (c) the exclusion of all methods from consideration that rely on the prior knowledge of range between the target and sensor to perform their functions, since in the assumed scenario the target's spatial scale in the image is expected to vary from the nadir viewing perspective.

So, a successful algorithm suite must be able to produce results that are invariant to the target's kinematic state, range, viewing perspective, and varying background clutter; also, be able to include on demand in its dictionary previously unknown targets. A key advantage using HS imagery as input in this context is that, since each pixel in the frame is represented by a vector (spectral profile), partial obscuration of targets and varying range would be naturally handled in most cases, depending on the algorithmic strategy.

Using real VNIR HS imagery, I propose an algorithm suite that focuses on different aspects of the target through three serial stage modules: (i) biometric based human skin detection, (ii) Bayesian Lasso inference [10] using all bands, and (iii) RGB quantization [11] using three bands.

The method in (i) consists of three steps: Step 1, reflectance retrieval; Step 2, exploitation of absorption wavelength line at 577 nanometers, due to oxygenated hemoglobin in blood near the surface of skin [12]; and Step 3, matched filtering on candidate patches in the input imagery that successfully passed Step 2, using as input all of the available bands in a spectral average representation of human skin. Step 3 is only applied to patches in the imagery showing evidence of human skin (Step 2 output).

The final output of method (i) serves as an initial spatial marker (*focus of attention*) for the remainder methods of the suite, whose action may start from having the user, observing a sequence of HS imagery and acting upon the receipt of intelligence information from an independent source(s), initiate the algorithm by first taking a spectral sample from one or more specific materials worn by the human target (e.g., hat, jacket, pants, shirt), and then triggering the suite to automatically search and find the specific material type(s) in the neighborhood of detected pixels representing human skin in the imagery. Results from methods (i), (ii), and (iii) are jointly considered as cumulative evidence of the target being present in the scene, promising to significantly reduce the false alarm rate. Real HS imagery is used to evaluate performance of the proposed cumulative evidence strategy.

The paper is organized as follows: Section II describes the data used for the analysis; Section III presents in more detail the proposed algorithm suite, Section IV discusses example results, and Section V concludes the paper.

II. DATA

The primary equipment used for the data collection included the Surface Optics Corporation (SOC), model SOC730VS. The SOC730VS is a hyperspectral imager with 240 bands covering the Vis and NIR spectral range from 500 to 1000 nm in wavelength. The SOC730VS is a *pushbroom* hyperspectral imager consisting of a diffraction based hyperspectral line imager with a computer controlled scan mirror on the front end. A BSI S9011 portable workstation equipped a cameralink interface and RS232 serial port was used for control and data acquisition. The hyperspectral data were acquired through the cameralink

interface, and the scan mirror was controlled using the RS232 serial interface. The SOC730 software was used for data acquisition and control, see Reference [13] for additional details and specifications of SOC730.

The data collection was held on December 2010 in the state of Maryland, USA, on the roof of a locally standard concrete building. The SOC hyperspectral imager was located at Position A, see Fig. 1, as two human subjects stood side by side at Positions B through H for approximately 1 min per acquisition time, while trying to minimize their motion. The range from the imager to the human subjects is labeled in Fig. 1 as B (50 ft or 15 m), C (100 ft or 31 m), D (150 ft or 46 m), E (200 ft or 61 m), F (250 ft or 76), G (300 ft or 91), and H (370 ft or 113 m).
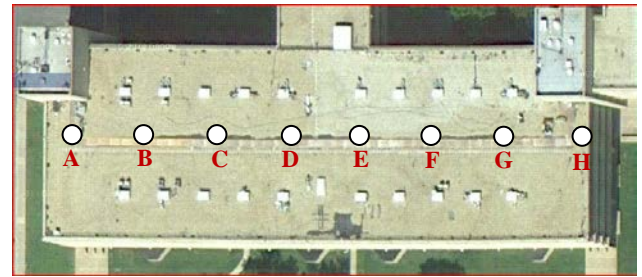

Fig. 1 Building rooftop and ground truth information, where a proof of principle experimentation in the VNIR was held using two human subjects standing at 50 ft (B), 100 ft (C), 150 ft (D), 200 ft (E), 250 ft (F), 300 ft (G), and 370 ft (H) from sensor location (A)

RGB images of the human subjects are shown in Fig. 2, where the subject standing in the left-hand side in each image has a skin type categorized as Type III (White to Olive) in the *Fitzpatrick Scale* [14] and the subject in the right-hand side has a skin type closer to Type I (Very Fair). (The Fitzpatrick Scale is used to describe skin color and its sensitivity to ultra violet radiation, where skin types range according to the following designation: Type I [Very Fair], Type II [Fair], Type III [White to Olive], Type IV [Brown], Type V [Dark Brown], and Type VI [Black].)

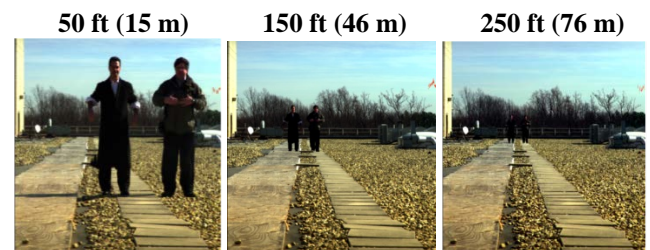| **50 ft (15 m)**	**150 ft (46 m)**	**250 ft (76 m)**

Fig. 2 Samples of scene consisting of two human subjects standing side by side at different ranges, they represent two different skin types in the Fitzpatrick Scale: Type III (White to Olive [left-hand side in each image]) and Type I (Very Fair [right-hand side])

The orange flag (upper right-hand side) and yellow pebbles across the scene are potential confusers for algorithms designed to exploit specific human skin spectral features in the VNIR; these particular scenery features prompted us to choose this location for data collection.

The human subject standing on the left (Type III skin) had the following skin areas exposed: the head, neck, hands, and forearms. The human subject standing on the right (Type I skin) had the following skin areas exposed: the head, part of the neck, and the hands. The subject on the right

wore a long dark blue winter coat, black pants, black shoes and white shirt, while the subject on the right wore a green coat, green pants, and brown shoes.

Also of particular interest are the presences of an orange-colored flag, which can be observed in the upper right-hand side of the RGB images in Fig. 2, and a multitude of pebbles in the ground featuring a color that is similar to the color of certain skin types. Those objects (flag and pebbles) in the scene are potential *confusers* for algorithms designed to exploit specific human skin spectral features in the VNIR. Such an algorithm is designed to automatically detect the presence of human skin in the imagery based on skin biometrics. The location of this data collection was in fact chosen because it provided those potential skin confusers naturally scattered across the scene.

During the data collection, the visibility was 10 miles with a 50% cloud cover consisting of thin and light clouds. The relative humidity was 30%. There had been no precipitation for previous 24 h up to and including the data collection period. Sunlight was shining on subjects and shadows were visible. The temperature was 54 °F (12°C) with average wind speed of 24 km/h, yielding a wind chill of 50 °F (10°C).

III. ALGORITHM STRATEGY

The overarching algorithmic strategy is shown in Fig. 3. The process starts when a spectral cue from a reference HS frame (Frame 1) is made available to the algorithm suite, $\mathbf{X}_{n \times B}^{(frame\,1)}$, where n is the number of spectra with B wavelength components.
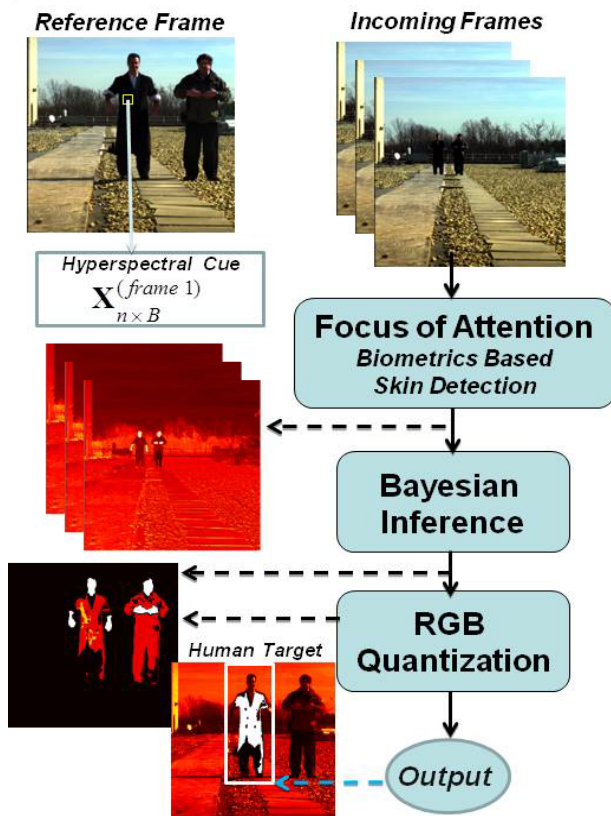

Fig. 3 Algorithmic suite

The goal is to use this spectral cue as the only reference in a test that is designed to detect the presence of the human target in incoming frames. The first stage, *focus of attention*, uses a particular absorption wavelength and two other neighboring wavelengths to automatically detect the spatial locations of human skin in the scene; Fig. 3 shows examples of this stage's output as *white* colored blubs spatially located where skin is potentially found (the method's details will be discussed shortly). A Bayesian Lasso method follows the first stage by inferring about the entire data cube (sparse approximation), using the inference results to test whether the sparse representation of $\mathbf{X}_{n \times B}^{(frame\,1)}$ matches those of similar material types in the scene. Fig. 3 shows an example of the Bayesian approach's output as clusters of *red* colored pixels, where in this particular case the spatial locations of both individual's clothes are highlighted. To mitigate potential false alarms, only the spatial areas embraced by red and white pixels are considered for further testing. The final stage tests whether the RGB color of a portion of or the entire object represented by red pixels in the Bayesian stage's output in Fig. 3 corresponds to the approximate color of $\mathbf{X}_{n \times B}^{(frame\,1)}$, using three RGB wavelength bands. Fig. 3 shows an example of the RGB quantization stage's output as *orange* colored pixels, emphasizing the output intersections between both stages (Bayesian and RGB quantization). Using the output results from all three stages, the algorithm displays the most likely spatial location where the human target is represented in the tested frame, see Fig. 3. Additional details of each stage are discussed next.

A. Biometric Based Human Skin Detection

The algorithmic approach proposed herein for human skin detection consists of three steps: Step 1, reflectance retrieval, using as input an uncalibrated VNIR hyperspectral data cube; Step 2, exploitation of the absorption wavelength line at 577 nm due to oxygenated hemoglobin in blood near the surface of skin; and Step 3, matched filter, using as input all of the available bands of a sample in the form of spectral average representation of human skin. Additional details follow.

Step 1: This step applies the in-scene QUAC (quick atmospheric correction) algorithm [15] for atmospheric correction and reflectance retrieval. QUAC is capable of autonomously retrieving an estimate of material reflectance from raw or uncalibrated hyperspectral data cubes at the expense of allowing an up to 30% of precision error. So, speed and independence of radiometric calibration requirements for the data are compromised by a potentially large estimation error.

Step 2: The method used in this step is based on analysis of previous spectral data experimentations (see, for instance, [12]), which point out the existence of a characteristic in the human skin spectral signature in the visible portion of the electromagnetic spectrum. The absorption feature owes to the presence of oxygenated hemoglobin in the blood near skin surface, whose overall effect produces a spectral signature that is independent of pose (front or profile) and generally different than those of background materials. The

skin spectral characteristic is the absorption near the 577-nm wavelength, where skin reflectance values as a function of wavelength are both accurately predictable and precise (Fig. 4).

The absorption line near the 577-nm band can be exploited in many ways. Using reflectance values at corresponding wavelengths $r(\cdot)$, I found through empirical means that the signal to noise ratio in the final output surface is significantly improved for the presence of human skin by applying the following metric:

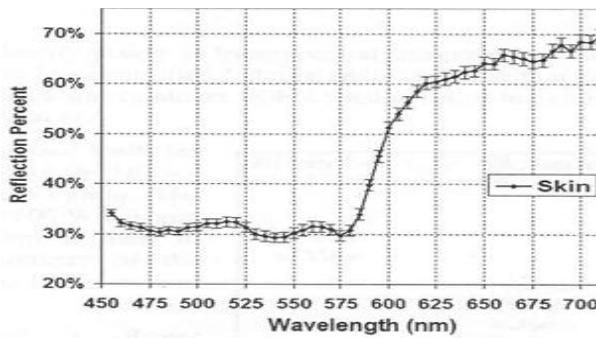$$Skin = \frac{[r(577\text{nm}+83\text{nm})]-r(577\text{nm})}{[r(577\text{nm}+43\text{nm})]} \quad . \tag{1}$$

Fig. 4 Absorption line at 577 nm due to oxygenated hemoglobin in blood near skin surface; narrow error bars (based on two standard deviations) ensure high precision of skin reflectance values

Using (1) to process the hyperspectral imagery shown in Fig. 2, the output surfaces shown in Fig. 5 (top row) were produced, where high values using (1) correspond to a high confidence level on the presence of human skin in the scene. The color map used to display the output surfaces in Fig. 5 (top row) shows white pixels as the most likely spatial locations in the imagery where skin is probably present in the scene. The artificial color ranges from yellow (representing other material types in the scene of characteristics closer to human skin), brown (less similar to skin), to black (most distinct from skin). Closer observation of Fig. 5 reveals that the orange colored flag (shown at the upper far right of each result) and the pebbles on the ground represent false alarms, thus, in this case, yielding a relatively high false alarm rate.

Step 3: In order to eliminate the nuisance false alarms produced in Step 2, the third step applies the SAM (spectral angle mapper) algorithm [1] using all of the bands of a previously available representation of skin to all skin-candidate patches in the imagery that successfully passed Step 2 screening. In particular, as a reference for SAM, the spectral sample average was used (computed using nine spectra) from the right hand of the human subject whose skin is Type III (subject at the left-hand side in each surface in Fig. 2). SAM is applied to contrast the reference spectrum and spectral averages that are computed using a 3x3 search window on all white pixel regions produced by Step 2. The same spectral reference was used independently of the range between the subjects and the sensor. Examples of Step 3 output surfaces are shown in Fig. 5 (bottom row) for ranges

50, 150, and 300 ft (or 15, 46, and 76 m, respectively). The color map shown in Fig. 5 (bottom row) is the same one used for the Step 2 output surfaces shown in Fig. 5 (top row). Notice that all of the false alarms—the waving flag and pebbles on the ground—are eliminated.
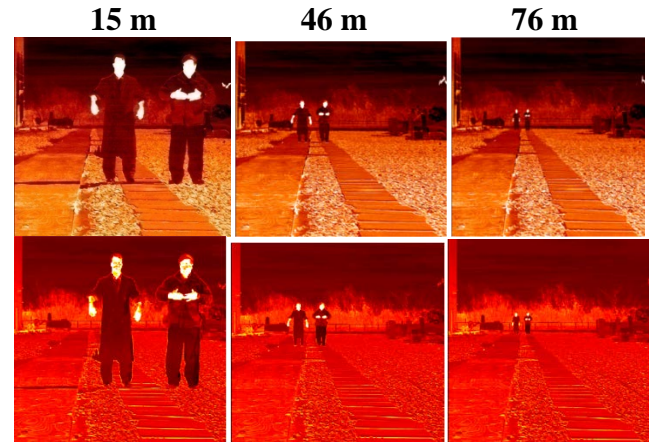
Fig. 5 Output surfaces resulted by using Steps 1-2 (top row) and Steps 1-3 (bottom row) for the scene shown in fig. 2. Notice in particular that the false alarms (pebbles on the ground and waiving flag) shown after Steps 1 and 2 are all eliminated, while the spatial locations of human skin evidence are preserved, according to the available ground truth. Similar performance for the other ranges

It is worth noting that false alarms produced by Step 2 may also have been due to errors produced by the reflectance retrieval algorithm used in Step 1. After all, QUAC is known to produce estimated reflectance within plus or minus 30% of actual values. But for robustness, we deliberately chose to use QUAC because it does not require the sensor to be well calibrated, which may be an important feature since many of these sensors are not properly calibrated when they are deployed to the field.

Although not presented in this paper, the results produced by Step 3 using as reference a spectral sample of subject skin Type III's right hand are comparable to results using skin spectral samples from elsewhere, e.g., forehead of subject skin Type III, forehead of subject skin Type I, left hand of subject skin Type I. Also, the results produced by a reduced algorithm suite (Step 1 and Step 3) produced a significant high number of false alarms whose spatial locations were mutually exclusive compared to the false alarms produced by an alternative reduced algorithm suite (Step 1 and Step 2). These outcomes motivated us to use the proposed three-step algorithm suite (Step 1, Step 2, and Step 3), as a baseline approach for human skin detection in the VNIR, potentially using as input uncalibrated HS imagery.

As mentioned earlier, the algorithmic approach presented in this section is intended to serve as a *focus of attention* in a multistage human target detection/tracking framework that may start from having a user, acting upon the receipt of intelligence information from an independent source(s), initiating the detection/tracking process by first taking a spectral sample from a specific material on the human target (e.g., hat, jacket), using as input remote sensing hyperspectral imagery, and then triggering an algorithm suite to automatically search and find the specific

material type (material target) near the spatial locations of human skin shown in consecutive imagery frames.

It is worth noting that if a hyperspectral-imagery-based algorithm suite can be developed to effectively accomplish the overarching task of human target detection from static imagery frames, then non-kinematic based tracking could be possible whether the target(s) is moving, stationary, in proximity to other objects, or partially obscured. Those challenges are examples of ambiguity problems, which are known for significantly degrading the performance of conventional tracking systems, where motion feature is a prime. In addition, the difficult subtasks of image geo-rectification and frame-to-frame registration would no longer be required, since the non-kinematic-based tracking approach would work independently on each frame at a time.

To complete the algorithm suite, two additional stages (Bayesian Lasso and RGB quantization) are proposed.

B. Bayesian Lasso Inference

Assume a HS image is measured and the *i*th block of data is cued by the user (see Fig. 3) and denoted $\mathbf{X}_i \in \Re^{n_x \times n_y \times B}$, where n_x and n_y represent the number of pixels in the two spatial dimensions, and B represents the number of sensor wavelengths. The model is fit for spectra for which data are available, and based on an inferred model (discussed further below) the missing values, if any, are imputed. For a given image we assume a set of vectors, $\{\mathbf{x}_i\}_{i=1,N}$, manifested by potentially considering all possible set of (possibly overlapping) blocks. In the subsequent discussion, the $\mathbf{X}_i$ will be assumed represented as an *unwrapped* vector $\mathbf{X}_i \in \Re^P$, with $P = n_x \cdot n_y \cdot B$

The factor model for each $\mathbf{X}_i$ is represented as

$$\mathbf{X}_i = \mathbf{D}\mathbf{s}_i + \boldsymbol{\varepsilon}_i \qquad (2)$$

where $\mathbf{D} \in \Re^{P \times K}$ has columns that define dictionary elements, $\mathbf{s}_i \in \Re^K$, and $\boldsymbol{\varepsilon}_i \in \Re^P$ represents noise (or model residual). Note that the dictionary $\mathbf{D}$ is shared across all vector $\{\mathbf{X}_i\}_{i=1,N}$, and the factor score $\boldsymbol{\varepsilon}_i$ is meant to be a sparse, and therefore only a subset of the dictionary elements (columns) are used to represent any particular $\mathbf{X}_i$. My objective is to infer the dictionary $\mathbf{D}$ based upon all $\{\mathbf{X}_i\}_{i=1,N}$, and the number of employed dictionary elements (used columns of $\mathbf{D}$ for representation of $\{\mathbf{X}_i\}_{i=1,N}$) is anticipated to be small relative to N (and small relative to P for large B). Once $\mathbf{D}$ is so learned, it may be used via the model to impute missing data, and the $\boldsymbol{\varepsilon}_i$ may be subtracted out, to remove noise.

A Bayesian setting for such a model places priors on the columns of $\mathbf{D}$, on the sparse vectors $\mathbf{s}_i$, and on the noise $\boldsymbol{\varepsilon}_i$. For the prior of $\boldsymbol{\varepsilon}_i$, I assume the *j*th component of $\boldsymbol{\varepsilon}_i$ may be drawn from the prior

$$\varepsilon_{ij} \sim N(0, \alpha^{-1}), \quad \alpha \sim \text{Gamma}(a_0, b_0) \qquad (3)$$

where the gamma probability density function (PDF) is represented by

$$\text{Gamma}(\alpha; a_0, b_0) = c\alpha^{a_0-1}\exp(-b_0\alpha), \qquad (4)$$

with $c = \beta^{a_0}/\Gamma(a_0)$, where $\Gamma(\cdot)$ is a gamma function; the gamma PDF is a *conjugate* prior for α, in that given observed data drawn from associated Gaussian distribution, the posterior of α, is also gamma distributed, with updated parameters (a_0, b_0). Note that such a prior is on the *marginal* probability for each component of noise, while the estimated posterior distribution does not assume the noise components are independent, and the full noise statistics are inferred. An important aspect of using such Bayesian constructs is that the noise statistics may be inferred (in terms of a posterior distribution on α) and need not be known *a priori*.

To impose sparsity on the factor score $\mathbf{s}_i$, I consider the Bayesian Lasso model [10]. In this model, I utilize the relationship

$$\frac{\sqrt{\gamma\alpha}}{2}\exp\left(-\sqrt{\gamma\alpha}|s|\right) = \int_0^\infty N\left(s; 0, (\alpha\xi)^{-1}\right)\text{InvGa}(\xi; 1, \gamma/2)d\xi \qquad (5)$$

where $\text{InvGa}(\cdot)$ represents the inverse-gamma distribution,

with $\text{InvGa}(\xi; a, b) = \dfrac{b^a}{\Gamma(a)}\xi^{-a-1}\exp(-b/\xi)$.

Assuming for a moment that the dictionary $\mathbf{D}$ is known, a draw of the data block $\mathbf{X}_i$ may be represented by

$$
\begin{aligned}
\mathbf{X}_i &\sim N\left(\mathbf{D}\mathbf{s}_i, \alpha^{-1}\mathbf{I}_p\right) \\
s_{ik} &\sim N\left(0, \alpha^{-1}\xi_{ik}^{-1}\right) \\
\alpha &\sim \text{Gamma}(a_0, b_0) \\
\xi_{ik} &\sim \text{InvGa}(1, \gamma_{ik}/2) \\
\gamma_{ik} &\sim \text{Gamma}(a_1, b_1)
\end{aligned} \qquad (6)
$$

where $\mathbf{I}_P$ is the $P \times P$ identity matrix.

In the context of HS data, further prior information should be exploited for modeling the *k*th column of $\mathbf{D}$, $\mathbf{d}_k$. Specifically, since signatures of materials are smooth functions of wavelength, this prior knowledge can be explicitly exploited rather than drawing the component of $\mathbf{d}_k$. i.i.d. from a normal distribution. Here, instead, $\mathbf{d}_k$ is drawn from a Gaussian *process* (GP).

For the GP construct, let $\lambda_1, \cdots, \lambda_B$ represent the sensor wavelength in increasing order. I wish to impose that for a given spatial location, the correlation between the signals at λ_j and $\lambda_{j'}$ increases with decreasing $|\lambda_j - \lambda_{j'}|$. The GP is a natural way to do this. Specifically, for each spatial location the wavelength-dependent component of each $\mathbf{d}_k$ is drawn from $N(0, \Sigma)$, where $\Sigma(j, j') = \Sigma(j', j) \geq \mathbf{0}$ represents the correlation between the signal at wavelengths λ_j and $\lambda_{j'}$. As customary in GP analysis, I assume the covariance matrix to have the form

$$\Sigma(j, j') = \zeta_1 \exp\left(-|j - j'|/\zeta_2\right) \qquad (7)$$

Separate gamma priors may be placed on both ζ_1 and ζ_2, although in this paper I simply set ζ_2 to promote a high probability of smoothness between consecutive wavelengths; a gamma prior is placed on ζ_1, allowing inference of an approximate posterior distribution on this parameter. As known, the GP construction does not require uniform sampling of wavelength, and once inference is performed using available data, it may be used to impute signal values at any other wavelengths (to infer the image at wavelength for which no data are measured). Using the Euclidean Distance (L_2), the resulting factor scores $\mathbf{s}_i$ from the local data inference are used as the discriminant feature to determine whether $\mathbf{X}_i$ corresponds to the target material type cued by the user, see Fig. 3. Details on parameter initialization are discussed in Section IV.

C. RGB Color Quantization

Assuming that variations in brightness can be ignored, potential false alarms produced by the Bayesian Lasso may be eliminated if they do not have the target material's RGB color. The well-known $L \cdot a \cdot b$ color space enables [11] one to quantify visual differences among distinct colors; hence, this method is applied here. The $L \cdot a \cdot b$ color space is derived from the so-called CIE XYZ tristimulus values [11]. The $L \cdot a \cdot b$ space consists of a luminosity layer L; a chromaticity layer a, indicating where color falls along the red-green axis, and a chromaticity layer b, indicating where the color falls along the blue-yellow axis. Since all of the color information is in the a and b layers, I use the K-means method [17] to cluster the objects using L_2. Objects, in this context, are clustered pixels with a and b values.

IV. RESULTS

The results presented in this section are based on analysis of the HS data cubes described in Section II. Details on the algorithm's implementation and performance results follow.

For skin detection, the metric in (1) was directly applied to the reflectance data in the corresponding wavelengths, as described in Subsection III-A. Concerning parameter settings in models described in Subsection III-B, parameters of the gamma priors on the precision of the noise were fixed at once to $a_0 = b_0 = 10^{-6}$ for all of the results. For the shrinkage model, corresponding parameters were fixed to $a_1 = b_1 = 10^{-6}$. In all experiments, the truncation level for the shrinkage model was set at $K = 128$. For the Gaussian process, the gamma prior was also set having both parameters equal to 10^{-6}. The shrinkage factor model was implemented with Gibbs sampling, with analytic update equations. For the results presented in here, I employed 100 burn-in iterations and 100 collection samples (a 10×10 window size was used for the cue, see Fig. 3); while this

number of samples is clearly insufficient to accurately estimate the full posterior distribution on all model parameters, it has in practice proven sufficient for estimation of mean parameters and the associated mean spectral representation of the different material types in the scene. For the RGB quantization, I used the following three bands to map RGB data to the $L \cdot a \cdot b$ color space: Red band (~0.680μm), Green band (~0.560μm), and Blue band (~0.465μm).

The methods of Section III were independently applied to each one of the three example HS data cubes shown in Fig. 3 (depicted as band averages), and their output results were jointly used to automatically detect the human target in each data cube. The target, in this example, is the material type featured in the winter coat of a person of interest at a particular time interval of interest. So, given a cued HS sample of the coat, the goal is to automatically detect and isolate the location of the human target in all of the data frames. The target material was cued $\mathbf{X}_i^{(frame\,1)}$ during a period of time when the human target stood at a range of 50 ft from the sensor; the small white box on the coat in Fig. 6 (top left) approximately shows the cue's spatial location. Fig. 6 (right column) shows the joint output results, using $\mathbf{X}_i^{(frame\,1)}$ as the only reference spectra for all data cubes.
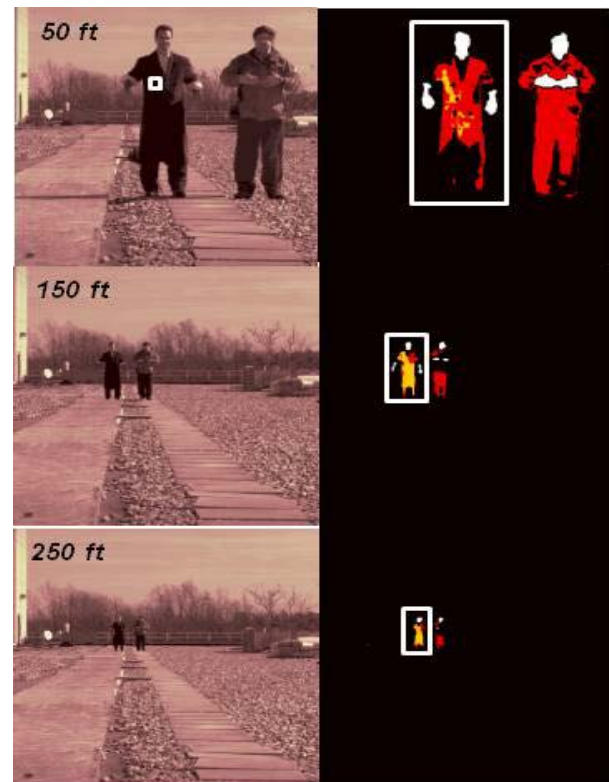

Fig. 6 Range invariant human target detection

The *white* pixel blubs (ignore the white boxes for now) in the output surfaces show the potential locations where human skin is present in the scene, according to the skin biometrics method described earlier; notice that the detection of hands and faces of both individuals is range invariant. The *red* pixel clusters in Fig. 6 (right column) show the locations where the coat material is potentially

present in the scene, according to the Bayesian Lasso approach; notice that the clothing material of the *no-target* human was also detected as false alarms (red pixels). Potential false alarms spatially located too distant from skin pixels (e.g., pixel distances > 10% of the image length) are automatically discarded at this stage. The *orange* pixel clusters in Fig. 6 (right column) show the locations where potentially similar or different material types have the same RGB color, according to the $L \cdot a \cdot b$ color quantization method; notice that in this example only a portion of the target material show similar color; however, in different scenarios, it is conceivable that distinct material types of similar color could be present elsewhere in the scene but be disassociated from any person in the scene; in those potential cases, they also would have been discarded. Finally, using the resulting cumulative evidences (i.e., the intersection of results produced by the Bayesian Lasso and RGB quantization methods spatially near human skin pixels), the algorithm suite places a *white* box, see Fig. 6 (right column), in the most likely location where the human target is present in the image. The performance depicted in Fig. 6 is invariant to range; comparable performance was observed for the other intermediate ranges, although they are not shown in here.

We emphasize the fact that the novelty presented in this paper is the proposed use of an algorithm framework that aims at significantly reducing the false alarm rate by fusing results of individual methods. These methods, when employed individually, normally produce a high false alarm rate addressing the operational scenario discussed herein. For instance, given the premises of the problem and goals of the effort, one could use the Support Vector Data Description (SVDD) method [18], and a kernel function [8], to automatically model the manually cued HS sample in a higher dimensional space than that of the native dimension of given data, where in that higher dimensional space the radius of the described circle surrounding the transformed reference sample is minimized relative to some other higher dimensional space. So in order to quantifiably compare the proposed algorithmic approach to other candidate methods, we chose to generate Receiver's Operating Characteristic (ROC) curves of the individual methods used in the proposed framework and include two additional methods: kernel SVDD and RX algorithm [19].

The RX algorithm is used in this context as a matched-filter based target detector, instead of the usual anomaly detector, where in the context of target detection the multivariate mean and covariance are estimated using the cued HS sample as the known target class. The RX test, thus, is reduced to checking the distance between each spectrum (in the data cube being tested) and the estimated reference target mean, and having this distance normalized by the estimated variability of the reference target sample. If the RX approach work as advertised, target spectra would show normalized low distance values compared to normalized distances produced by non-target spectra.

Since other candidate methods that could also be used to address the problem described in this paper fall under the class of kinematic based approaches, they are not considered

here for comparison, because features based on target motion are not a viable option; recall that some of the requirements stated earlier imply that targets may be stationary, accelerating, disaccelerating, or obscured for short or prolonged periods of time.

Figure 7 shows the cumulative ROC curves produced by the output of five algorithms testing the three data cubes shown in Fig. 6. The algorithms include the three individual methods of the proposed framework, denoted in Fig. 7 as Skin (focus of attention in Fig. 3), Bayesian (Bayesian Inference in Fig. 3), and RGB (RGB in Fig. 3), in addition to SVDD (kernel SVDD) and RX.
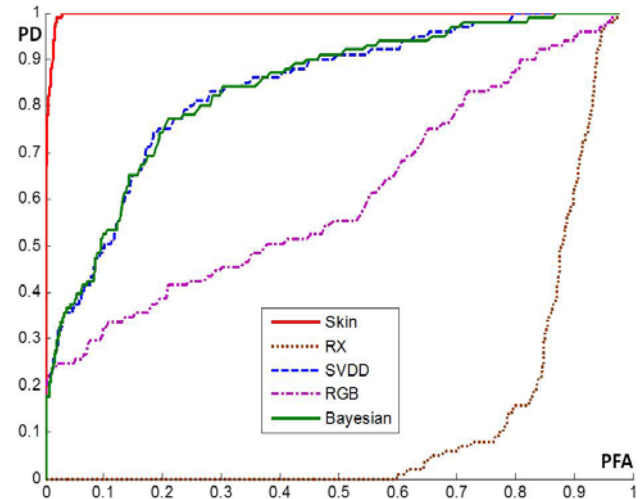

Fig. 7 Comparative ROC curve performances among the three methods used in the proposed framework (Skin, Bayesian, RGB) and the well-known kernel SVDD and the RX algorithm (functioning here as a matched filter based target detector)

The fusion of detectors Skin, Bayesian LASSO, and RGB yielded a perfect performance (PD = 1, PFA = 0), since it was able to isolate the intended human target (see Fig. 6), independently of range and the mutually exclusive false alarms produced by the individual methods, which is not evident from merely observing their ROC curves. The *target* of the Skin detector is the genetic human skin; the *target* of the other detectors is the winter coat material worn by the human target.

Detection performance was measured using the ground truth information for the data cubes. We used the coordinates of all of the target spatial regions to represent the ground truth target set, call it *Target-Truth-50ft*, for data cube collected at the 50 ft range, *Target-Truth-150ft*, for data cube collected at the 150 ft range, and *Target-Truth-250ft*, for data cube collected at the 250 ft range. If we further denote the region outside the *Target-Truth-50ft* as *Clutter-Truth-50ft*, then the intersection between *Target-Truth-50ft* and *Clutter-Truth-50ft* is *zero* and the entire scene in data cube at the range 50 ft is the union of *Target-Truth-50ft* and *Clutter-Truth-50ft*. The same rationale is used to define *Clutter-Truth-150ft* and *Clutter-Truth-250ft*.

In Fig. 7, for a given decision threshold, the proportion of target detection (PD) was measured as the proportion between the total cumulative number of detected pixels belonging to *Target-Truth-50ft* (n_{50}), *Target-Truth-150ft* (n_{150}), and *Target-Truth-250ft* (n_{250}), over <u>all</u> pixels

belonging to *Target-Truth-50ft* (N_{50}), *Target-Truth-150ft* (N_{150}), and *Target-Truth-250ft* (N_{250}), i.e.,

$$PD = \frac{n_{50} + n_{150} + n_{250}}{N_{50} + N_{150} + N_{250}} \qquad (8)$$

In Fig. 7, the proportion of false alarms (*PFA*) was measured as the proportion between the total cumulative number n of detected pixels in *Clutter-Truth-50ft*, *Clutter-Truth-150ft*, and *Clutter-Truth-250ft*, over all pixels belonging to *Clutter-Truth-50ft*, *Clutter-Truth-150ft*, and *Clutter-Truth-250ft*, i.e.,

$$PFA = \frac{n}{N - (N_{50} + N_{150} + N_{250})} \qquad (9)$$

where N is the total number of pixels covering the entire spatial areas of data cube (50 ft), data cube (150 ft), and data cube (250 ft).

In general, the quality of a target detector can be readily assessed by noticing a key feature in the shape of its ROC curve: the closer the *knee* of a ROC curve is to the *PD* axis, the less sensitive the approach is to different decision thresholds. In other words, *PFA* does not change significantly as *PD* increases. In essence, an ideal ROC curve resembles a step function that starts at point $(PFA = 0, PD = 1)$. Note: the definition of *target* in Fig. 7 depends on the detector, i.e., for the Skin detector, the target is the genetic human skin; for detectors Bayesian, SVDD, RX, and RGB, the target is the winter coat worn by the human subject shown at the left hand side in Fig. 6.

Fig. 7 shows some interesting points worth highlighting. For instance, notice the obvious poor performance of a standard matched filter based approach (RX, being used here as a target detector); this performance would be significantly improved if the search using RX were restricted only to areas in the scene showing signs of human skin, where the bulk of the false alarms shown in Fig. 7 (RX) of up to 0.6 (PFA)—just before the first sign of the target is manifested—would be eliminated since these pixel locations fall significantly far away from pixel representing skin.

The benefit of the proposed framework can be so powerful that, by just using the Skin and RX detectors in a two-stage framework, the RX ROC curve would completely shift to the left, eliminating the flat line at zero PD between 0 and 0.6 in the PFA axis. This is possible because the ROC curve produced by the Skin detector essentially behaves as an ideal step function starting at zero PFA, working almost perfectly across the three data cubes. However, as mentioned earlier, the skin detector does not isolate the intended target (i.e., the human subject shown in the left hand side in Fig. 6) from other regions that also feature skin presence.

Regarding the kernel SVDD and Bayesian LASSO methods' performances in Fig. 7, one cannot ignore the striking resemblance between their ROC curves, which may suggest that, independently of the target feature being exploited for target detection, different kinds of advanced methods (e.g., kernel functions and support vector concepts, Bayesian LASSO, sparse based approach) may yield equivalent performance. (Attempting to rationalize the equivalent performance between kernel SVDD and Bayesian LASSO is beyond the scope of this paper, but we plan on addressing it in the future.)

Last but not least observation from Fig. 7 is what it seems a mediocre performance by the RGB detector, since it only uses three bands, further emphasizing the importance of accepting the fact that, since each independent method is pragmatically flawed (i.e., they will produce false alarms in order to detect the target). The best possible approach then is to find a number of methods that will exploit orthogonal attributes of HS objects, such that, at the end, these detectors will always jointly find the target while producing expected false alarms that are mutually exclusive among the employed methods. We conjecture that the proposed algorithmic strategy will *always* work significantly better than using any particular single method.

For instance, the results shown in Fig. 6 reflect a perfect performance (PD = 1, PFA = 0), since the fusion of the proposed three methods (Skin, Bayesian, and RGB) was able to isolate the intended human target, independently of range and the false alarms produced by the individual methods, as shown by their ROC curves in Fig. 7. Using only results from testing the reference data cube (range 50 ft), the criteria used to obtain thresholds for results shown in Fig. 6 are as follows: (i) find the threshold that can yield a chosen PD = 0.85 (PFA = 0.00) using the Skin detector, (ii) find the threshold that can yield a chosen PD = 0.70 (PFA = 0.19) using the Bayesian detector, and (iii) find the threshold that can yield a chosen PD = 0.01 (PFA = 0.00) using the RGB detector. A point that must be reemphasized is that these thresholds, based solely on a single data cube (range 50 ft), were fixed at once to test the other data cubes (ranges: 150 ft and 250 ft), and the same thresholds also yielded a perfect performance (PD = 1, PFA = 0) on the tested data cubes.

V. CONCLUSION

This paper introduced an algorithm suite that exploits three independent characteristics of a human target in VNIR HS data: (i) human skin, (ii) Bayesian Lasso inference on a specific material worn by the target, and (iii) the material's RGB color. Detections from the Gaussian Lasso model are only considered if they are spatially located near the presence of human skin. Additional constraint is imposed by considering the target material's RGB color relative to other material types in the scene. The strategy yielded promising results using real HS data. Note that I am not claiming the possibility of human target identification (ID) through biometrics or some other ID feature. Instead, in this context, the uniqueness of a human target is assumed to exist through the material's spectral signature worn by the target relative to other material types in the scene, during the particular time interval of interest. Interestingly, if the human target detection task can be reliably and independently performed for each consecutive image frame,

the algorithm results may be interpreted as tracking, i.e., non-kinematic based target tracking, since motion feature is not utilized. Non-kinematic based target tracking is immune to the most difficult scenarios facing kinematic based trackers, e.g., target's sudden acceleration/deceleration, target in proximity to another moving object, short/long obscurations. As follow up, I plan on exploiting the shortwave infrared (SWIR) region of the spectrum, since SWIR frequencies are known for penetrating fog, haze, smoke, dust, and other obscurants caused by the atmosphere.

REFERENCES

[1] R. Schowengerdt, Remote Sensing, Models and Methods for Image Processing, 2nd Ed., Academic, San Diego, 1997.

[2] Subramanian S., Gat N., Subpixel Object Detection Using Hyperspectral Imaging for Search and Rescue Operations. *Automatic Target Recognition VIII* April 1998, *3371*(1), 216–225.

[3] Simard J., Mathieu P., Fournier G., Larochelle V. A., Range-Gated Intensified Spectrographic Imager: an Instrument for Active Hyperspectral Imaging. *Proceedings of Laser Radar Technology and Applications V* 2000, *4035*, 180–191.

[4] Simi C., Hill A., Kling H., Airborne Remote Spectrometry Support to Rescue Personnel at 'Ground Zero' After the World Trade Center Attack on September 11, 2001. *Proc of Imaging Spectrometry VIII*, 4816, 23–32, 2002.

[5] Yang M., D.J. Kriegman, and N. Ahuja. "Detecting faces in images: A survey". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1): 34-58, 2002.

[6] Pan Z., Healey G., Prasad M., Tromberg B., Hyperspectral Face Recognition under Variable Outdoor Illumination. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X* August 2004, *5425*, 520–529.

[7] Pan Z., Healey G., Prasad M., Tromberg B., Face Recognition in Hyperspectral Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* December 2003, *25*(12), 1552–1560.

[8] V.N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd Ed., Springer, New York, 2000.

[9] A. Gelb, *Applied Optimal Estimation*, MIT, Boston, 1979.

[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. Series B*, v. 58, pp. 267-288, 1994.

[11] T. Smith, J. Guild, "The C.I.E. colorimetric standards and their use," *Tran. Optical Society*, v. 33, pp. 73-134, 1931.

[12] Van Gemert, Jacques Sterenborg, "Skin Optics," *Trans. Biomedical Eng.*, IEEE, NY, 36(12), 1146–1154, 1989.

[13] Surface Optics Corp homepage http://www.surfaceoptics.com (accessed 14 March 2012).

[14] Wagner J. K., Jovel C., Norton H. L., Parra E. J., Shriver M. D., Comparing Quantitative Measures of Erythema, Pigmentation and Skin Response using Reflectometry. *Pigment Cell Research* 2002, (15), 379–384.

[15] Bernstein, L. S.; Adler-Golden, S. M.; Sundberg, R. L.; Ratkowski, A. J. Improved Reflectance Retrieval from Hyper- and Multi-spectral Imagery without Prior Scene or Sensor Information. *Proc. SPIE* Vol. 6362, Remote Sensing of Clouds and the Atmosphere XI, 63622P (Oct. 12, 2006).

[16] Shaw G., Manolakis D., Signal Processing for Hyperspectral Image Exploitation. *IEEE Signal Process. Mag.* Jan. 2002, *19* (1), 12–16.

[17] Duda R. O., Hart P. E., Stork D. G., Pattern Classification, 2nd Ed., John Wiley & Sons, New York (2001).

[18] Tax D. M. J., Duin R. P. W., "Support Vector Data Description," Machine Learning, vol. 54, No. 1, pp. 45-66, Jan. 2004.

[19] Yu X., Hoff L., Reed I., Chen A., Stotts L., "Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach," *IEEE Tran. Image Processing*, 6, pp. 143-156 (1997).

Dalton S. Rosario received the B.S. degree in electrical engineering from the California State University, Fresno, in 1990, the M.S. degree in electrical engineering from the Johns Hopkins University in 1996, and the Ph.D. degree in applied mathematics from the University of Maryland, College Park, in 2008. Since 1990, he has been with the U.S. Army Research Laboratory (ARL) studying the detection and classification of broadband and hyperspectral targets. During his tenure with ARL, he has served as team leader and technical lead for various projects, as a member of advisory panels of the U.S. Dept. of Energy funded university projects on hyperspectral image processing, and as the action officer for the U.S. Army RDECOM Sensors Technology Focus Team, in direct support of the ARL/SEDD Director. Dr. Rosario holds a U.S. patent in hyperspectral anomaly detection and a pending patent in hyperspectral target tracking. His research interests include mathematical statistics, multivariate image processing, detection theory, and the application of hyperspectral and polarimetric sensing technologies.